AEM
Journals.ASM.org

Culture-Independent Analysis of Aerosol Microbiology in a Metropolitan Subway System

Charles E. Robertson,[a] Laura K. Baumgartner,[a]* J. Kirk Harris,[b] Kristen L. Peterson,[a] Mark J. Stevens,[b] Daniel N. Frank,[c] Norman R. Pace[a]

Department of Molecular, Cellular, and Developmental Biology, University of Colorado, Boulder, Colorado, USA[a]; Department of Pediatrics, University of Colorado School of Medicine, Aurora, Colorado, USA[b]; Division of Infectious Diseases, University of Colorado School of Medicine, Aurora, Colorado, USA[c]

The goal of this study was to determine the composition and diversity of microorganisms associated with bioaerosols in a heavily trafficked metropolitan subway environment. We collected bioaerosols by fluid impingement on several New York City subway platforms and associated sites in three sampling sessions over a 1.5-year period. The types and quantities of aerosolized microorganisms were determined by culture-independent phylogenetic analysis of small-subunit rRNA gene sequences by using both Sanger (universal) and pyrosequencing (bacterial) technologies. Overall, the subway bacterial composition was relatively simple; only 26 taxonomic families made up ~75% of the sequences determined. The microbiology was more or less similar throughout the system and with time and was most similar to outdoor air, consistent with highly efficient air mixing in the system. Identifiable bacterial sequences indicated that the subway aerosol assemblage was composed of a mixture of genera and species characteristic of soil, environmental water, and human skin commensal bacteria. Eukaryotic diversity was mainly fungal, dominated by organisms of types associated with wood rot. Human skin bacterial species (at 99% rRNA sequence identity) included the *Staphylococcus* spp. *Staphylococcus epidermidis* (the most abundant and prevalent commensal of the human integument), *S. hominis*, *S. cohnii*, *S. caprae*, and *S. haemolyticus*, all well-documented human commensal bacteria. We encountered no organisms of public health concern. This study is the most extensive culture-independent survey of subway microbiota so far and puts in place pre-event information required for any bioterrorism surveillance activities or monitoring of the microbiological impact of recent subway flooding events.

The microbiological quality of the air we encounter daily, and depend upon absolutely, is a significant yet little-addressed societal concern. The regulatory focus on air quality has been on chemical and particulate materials, which are readily measured. However, there is comparatively little knowledge of the nature of the aerosolized microorganisms and microbial products that occur in different public settings and to which the public is exposed daily. Because many public places concentrate large numbers of humans and therefore may be key locales for the transmission of natural pathogens or deliberately released agents, understanding the microbial ecology of bioaerosols in public settings is critical for public health, occupational health, and biodefense. For instance, prior knowledge of the composition, sources, and temporal and spatial dynamics of bioaerosol microbes is essential for tracking pathogen dispersal in public settings.

One public arena through which large numbers of people pass daily is the municipal subway system. Subway facilities have been established by major metropolitan areas throughout the world. New York City (NYC) subways, for instance, had a ridership of 1.6 billion in 2011 (http://www.mta.info/nyct/facts/ridership/index.htm#atGlance_s). The huge number of people exposed to the subway environment underscores the importance of developing some understanding of the microbiological air quality of subway platforms.

The results of previous studies emphasize that airborne particulate materials in subways are different from what is found on city streets or in other indoor environments. This is particularly due to aerosolized metallic dust, which most likely is generated by the action of iron train wheels on tracks (1–6). Chillrud and colleagues examined the effects of NYC subway exposure on children (7) and subway workers (8, 9) and saw no health effects. Birenzvige et al. showed that the load of particulate matter in the air on subway platforms was correlated to the frequency of train traffic (10). The unusual particulate quality of the subway environment may indicate also that the microbial contents of subway aerosols differ from those detected in other indoor or outdoor environments.

Thus far, studies of aerosol microbiology in the subway environment have focused primarily on culture-dependent techniques, most notably, viable counts of bacteria and fungi, sometimes with biochemical or molecular identification of cultivars (2, 11–14). However, studies that focus on CFU likely identify only a small fraction of the actual microbial contents because most environmental microbes are not cultured by standard techniques (15). Consequently, there is little general perspective on the nature of microorganisms that might be encountered in subway air.

In order to gain an overview of the nature of the microbial load of a subway system independently of culture, we examined the contents of aerosols collected on several New York City subway

Received 31 January 2013 Accepted 26 March 2013
Published ahead of print 29 March 2013
Address correspondence to Norman R. Pace, norman.pace@colorado.edu.
* Present address: Laura K. Baumgartner, Front Range Community College, Longmont, Colorado, USA.
C.E.R. and L.K.B. contributed equally to this study.
Supplemental material for this article may be found at http://dx.doi.org/10.1128/AEM.00331-13.

doi:10.1128/AEM.00331-13

platforms and associated sites (terminal, park, and unused platform) during three sampling sessions over a 1.5-year period. The compositions of aerosolized microbial loads were determined by culture-independent phylogenetic analysis of small-subunit (SSU) rRNA gene sequences by both Sanger and 454 sequencing technologies. The longer Sanger sequences provide more accurate phylogenetic identification, whereas pyrosequences provide more comprehensive sampling of aerosol microbiology.

MATERIALS AND METHODS

Air sampler setup, cleaning, and blanks. Air was sampled with a custom-modified Omni 3000 fluid impinger (InnovaPrep LLC, Drexel, MO) modified to minimize tubing and allow full removal of tubing for cleaning and sterilization. Omni particle capture efficiency ranges from 40% for 0.5-μm particles to >90% for >3-μm particles (16). The device was chosen because of its relatively high capture rate of bacterium-sized particles; high rate of air passage (250 to 300 liters/min in this study), to minimize sampling time; and relatively quiet function, to minimize conspicuity during this public sampling campaign.

Empty sampling cartridges were sterilized by irradiation in a UV Stratalinker 1800 (Stratagene, La Jolla, CA) and then molded shut with chloroform. Cartridges were filled with a filter-sterilized impinging solution that consisted of phosphate-buffered saline (137 mM NaCl, 10 mM phosphate, 2.7 mM KCl, pH 7.4) and 0.005% Tween in diethyl pyrocarbonate (DEPC)-treated water. Blank cartridges were prepared at the same time as the sample cartridges, carried during sampling, and processed in the same manner as the air samples. None of the cartridge blanks produced amplifiable DNA. The average air-sampling rate was ~300 liters/min. During sampling, the liquid volume lost to evaporation was replaced with DEPC-treated water. The air sampler was cleaned between stations by replacing all of the tubing with sterile tubing and cleaning the contactor and ports with DEPC-treated water and alcohol (isopropanol wipe and then washing with 70% ethanol) and air dried. Cleaning blank samples were taken by filling the sampler (contactor and tubing) with 5 ml DEPC-treated water, which was allowed to sit in the sampler for 5 min and then extracted with a sterile syringe from the sampling port. The cleaning blanks were filtered and extracted in the same fashion as the air samples; none of the blanks produced amplifiable DNA.

DAPI counts. Aliquots of sample fluid were adjusted immediately following sampling to 4% formaldehyde for direct cell counts and held and shipped overnight on wet ice to the laboratory. Aerosolized iron in most subway air samples interfered with microscopic counts and was removed before counting; samples were vortexed gently for 30 s and then placed against a magnetic rack for ~1 min. Sample fluids were carefully removed with a pipette, and a known volume was filtered through a 0.2-μm black polycarbonate filter. The filter was placed on a slide, flooded with 10 μg/ml 4′,6-diamidino-2-phenylindole (DAPI) for 5 to 10 min, and washed twice for 2 min each time with 10 mM Tris-Cl–1 mM EDTA (pH 7.4), and slides were counted at a magnification of ×400.

DNA extraction, amplification, and sequencing (Sanger and 454). Most (~85%) of the sample fluid from each collection was filtered through 0.2-μm polycarbonate filters (Isopore; Millipore, Billerica, MA), which were then placed in sterile microcentrifuge tubes, immediately frozen in liquid nitrogen, and shipped frozen to the laboratory, where they were stored at −80°C until processed.

Samples were processed to remove particulate iron before DNA extraction by dissolving the filter in phenol, chloroform, and buffer B (200 mM NaCl, 200 mM Tris-Cl [pH 8.0], 20 mM EDTA, 5% SDS) during vortex mixing. Samples were centrifuged at 100 × *g* for 2 min and then placed against a strong magnet. All liquids were removed to a new sterile tube, leaving the magnetic iron behind. This was repeated, and the sample was then placed in a tube with zirconium beads for DNA extraction by bead beating with 2 volumes of buffer-saturated phenol and ethanol precipitation of nucleic acids.

Extracted genomic DNA was amplified for cloning by PCR with (nominally) universal SSU rRNA gene primers 515F and 1391R (17). PCRs were conducted at 94°C for 2 min, followed by 30 cycles of 94°C for 20 s, 52°C for 20 s, and 65°C for 1.5 min, followed by a 65°C elongation step for 10 min. Each 50-μl reaction mixture contained 10 μl Eppendorf 2.5× HotMasterMix (Eppendorf, New York, NY), 10 μl water, 0.05% bovine serum albumin (Sigma-Aldrich, St. Louis, MO), 100 ng of each oligonucleotide primer, and 1 to 5 ng of template DNA. Triplicate PCRs were conducted for each sample and pooled before purification with the Montage gel purification system (Millipore). In some (9/27) gel purifications of PCR products, 16S and 18S bands were processed and analyzed separately. For most of the universal libraries, a mixture of the two was processed. Some samples were not effective as templates and were not analyzed further. PCR-amplified rRNA genes were cloned with Topo-TA according to the manufacturer's instructions (Life Technologies, Carlsbad, CA), and Sanger sequencing was conducted on an Amersham MegaBACE 1000 capillary sequencer in accordance with the manufacturer's protocols.

DNA samples also were analyzed by pyrosequencing on a Roche 454 GS-FLX platform. DNAs were amplified in three independent reaction mixtures using barcoded primers (27F-338R) (18). Negative PCR controls for each primer were assayed in parallel and did not exhibit bands in agarose gels. The three independent reactions were pooled, and amplicons were confirmed by agarose gel electrophoresis. DNA contents of pools were normalized by using the SequalPrep Normalization Plate (Life Technologies), and equal amounts were mixed to construct the amplicon pool (19). The amplicon pool was concentrated by evaporation, size selected by electrophoresis on a 1.5% agarose gel (Tris-acetate-EDTA buffer), and gel purified by Montage kit (Millipore) prior to sequencing. Pyrosequencing was conducted in accordance with the manufacturer's protocols by using Roche 454 titanium chemistry.

Sequence analysis. Sanger sequences were quality filtered and assembled with XplorSeq (20). Raw pyrosequences were quality filtered and sorted into their respective barcoded libraries with BARTAB (21). Filtering for both Sanger and pyrosequence data removed nucleotides with mean Q values of <20 at the 5′ and 3′ ends and over a 10-nucleotide (nt) window, sequences with more than one ambiguous base were discarded, and all sequences with lengths of <200 nt were discarded. Infernal (22) and ChimeraSlayer (23) were used to screen bacterial sequences as described previously (24). Taxonomic classification of all sequences was done with the classifier functionality of standalone SINA by using the SILVA 108 nonredundant database as the reference (376,437 sequences that are the 99% cluster representatives of SILVA 108 [25]). Ecological statistics (such as the number of species observed [26]), pie charts, and heat maps were prepared with the Explicet software package (C.E.R., unpublished data; software available upon request).

Determination of potential sources of DNA was done by BLAST analysis of all pyrosequences against three separate databases composed of long sequences (≥1,200 nt) associated with the Human Skin Microbiome (HSM) study (27), sequences in SILVA 108 whose isolation source metadata tag contained the word "soil," and sequences in SILVA 108 whose isolation source metadata tag contained the word "water." To be considered a BLAST database match, pyrosequences were required to overlap the BLAST database hit sequence by a minimum of 95%, have a minimum bit score of 50, and be either 95 or 99% identical to the respective BLAST database hit sequence. We did not use tools that compare fingerprints of ecosystems with bioaerosol samples because air is an assemblage, not a specific microbiome.

Nucleotide sequence accession numbers. The Sanger sequences determined in this study have been deposited in GenBank under accession numbers JX394222 to JX397762, and the pyrosequences determined have been deposited in the Short Read Archive under accession number SRA055336 (BioProject accession number PRJNA169671).

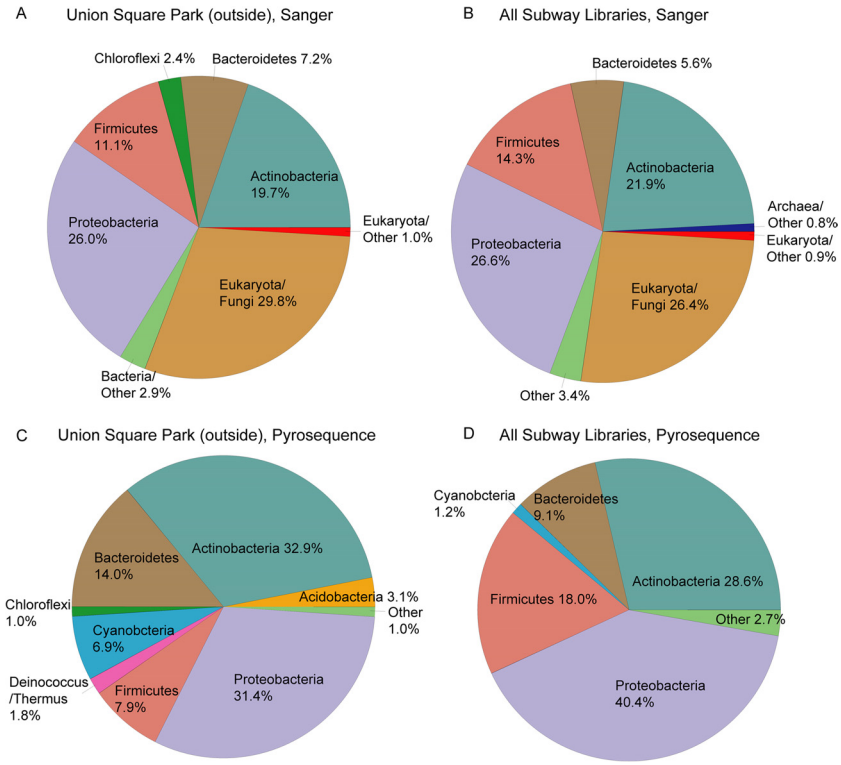

FIG 1 Phylum level distribution of aerosolized microbiota observed at NYC subway stations and Union Square Park (outside). Shown are the percent abundances of phyla observed among the Sanger sequences and the pyrosequences extracted from impinged air sampled at Union Square Park, a proxy for outdoor air (A and C), and eight subway stations (C and D). A and B, Sanger sequence data (bacteria, archaea, and eucarya); C and D, pyrosequence data (bacteria only).

RESULTS

The microbiological contents of nominally similar environmental samples tend to vary from site to site and over time. Variation is particularly expected in the subway setting, a scattered collection of potentially more or less sequestered spaces connected by kilometers of track tunnels. In order to explore both the spatial and temporal aspects of the microbiology of the NYC subway system, we collected air samples at seven subway stations (specified in the figures and located as shown in Fig. S1 in the supplemental material) and three adjacent sites in lower Manhattan over a 1.5-year period (2007 to 2008). Two 5- to 6-m^3 samples were collected at each subway station for each sampling time point, one from each end of the platform, with a metered, high-volume fluid impinger. The impinger was modified so that all internal surfaces could be exposed for cleaning or disposal following their use (see Materials and Methods).

Sampling parameters and other metadata are summarized in Table S1 in the supplemental material. Microbial cell counts (see Materials and Methods) of typically 1×10^4 to 4×10^4 cells/m^3 were observed throughout the system (average, 2.2×10^4 cells/m^3; see Table S1). These directly counted microbial loads were typical of outdoor air samples taken at the same time and at the low end of loads typically encountered in indoor environments such as occupied buildings (10^4 to 10^5 cells/m^3) (28–30). On the other hand, direct counts of aerosolized subway bacteria are much higher, ~100-fold higher, than CFU counts encountered in subway environments, which are typically several hundred CFU/m^3 in different subway systems (12–14). We see no clear patterns of variation in cell counts by season, station, or volume of local traffic (light versus heavy; see Table S1) in this sampling.

The phylogenetic distribution of aerosolized microbes in each specimen was determined by analysis of SSU rRNA gene sequences. Total nucleic acids were extracted from impinger fluids by a bead-beating protocol and processed for Sanger sequencing or pyrosequencing analysis (see Materials and Methods). Clone libraries for Sanger sequencing were constructed from PCR products obtained with the (nominally) universal primer set 515F-1391R; pyrosequencing libraries were developed with bacterium-specific barcoded primers targeting the V1-V2 region of the SSU rRNA gene. The longer Sanger sequences (700 to 800 nt in this case) provide for relatively accurate phylogenetic assignments and a three-domain census. The more numerous but shorter (~200-nt) pyrosequences provide a broader census assessment of bacteria. Overall, 3,541 Sanger sequences and 60,707 pyrosequences were determined and deposited in GenBank.

Phylum level diversity of bioaerosols. The phylum level diversities encountered among Sanger sequences (universal) and pyrosequences (bacteria only) were similar throughout the subway sites examined and similar to those of outdoor air samples (Union Square Park, as summarized in Fig. 1. Qualitatively similar subway and outdoor bacterial diversities were found in the Sanger sequence (Fig. 1A and B) and pyrosequence (Fig. 1C and D) analyses, although there was some quantitative variation. Few (<1%) of the sequences were identical to sequences in the databases, but many were related to database sequences at the genus or even species taxonomic level (below). On the basis of the results of PCR

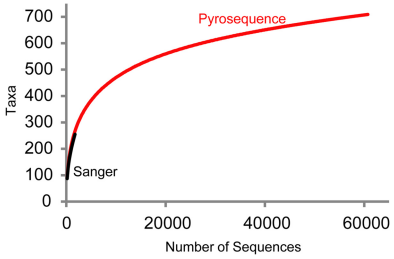

FIG 2 Comparison of the numbers of species observed in the Sanger and pyrosequence data sets. Shown are collector curves of the number of taxonomic classifications versus sequence sample size. Sanger sequence values are in black, and pyrosequence values are in red.

amplification with universal primers, bacterial (65%) and fungal (34%) sequences dominated the overall subway rRNA assemblage (Fig. 1B). Archaeal sequences, consisting of a few crenarchaeotes and methanogens with no specific known relatives, were rarely encountered (<1% of the total). Collectively, the main bacterial sequence diversity observed in the subway environment was remarkably simple, mainly comprising only 4 of the ~100 known bacterial phyla, *Actinobacteria*, *Firmicutes*, *Bacteroidetes*, and *Proteobacteria*, with only minor contributions from other phyla. Only a few sequences fell into candidate phyla (OP10 and TM7) with no cultured representation. For a full, abundance-ranked list of the bacterial taxa detected in this study, see Table S2 in the supplemental material.

Although simple at the phylum level, the diversity of subway air samples was more complex at lower taxonomic levels. Comparison of taxon collector curves (Fig. 2) generated from pyrosequence and Sanger data sets indicates that the Sanger sequence set, at 700 to 800 nt in length, substantially undersamples the microbial diversity relative to the pyrosequencing data set (~200-nt length), although the most abundant sequences are expected to occur in the Sanger data set (31). However, even the pyrosequencing data set did not exhaustively sample the microbes present in the system as a whole, as indicated by the continuous increase in the number of species observed with ongoing sampling of sequences (Fig. 2). This is a common theme in environmental sequence surveys; the pool of rare sequences, the "rare biosphere," is always extensive (32). Nonetheless, pyrosequencing was sufficiently extensive to cover most of the microbial diversity encountered. Consequently, we focus on the pyrosequences for site-to-site statistical comparisons.

Bacterial diversity. Despite the considerable diversity of rRNA sequences encountered in subway air samples, only 26 family level taxa made up most (~75%) of the sequence types observed; their distributions in the different stations, abundance ranked by heat map, are shown in Fig. 3. (Table S1 in the supplemental material lists all of the taxa encountered to 0.01% of the total.) At this taxonomic level of analysis, the makeup of the subway microbiology samples was generally similar to that of the outdoor samples (Union Square Park) in an area with considerable human traffic. Most of the abundant taxa occurred throughout the sample set at

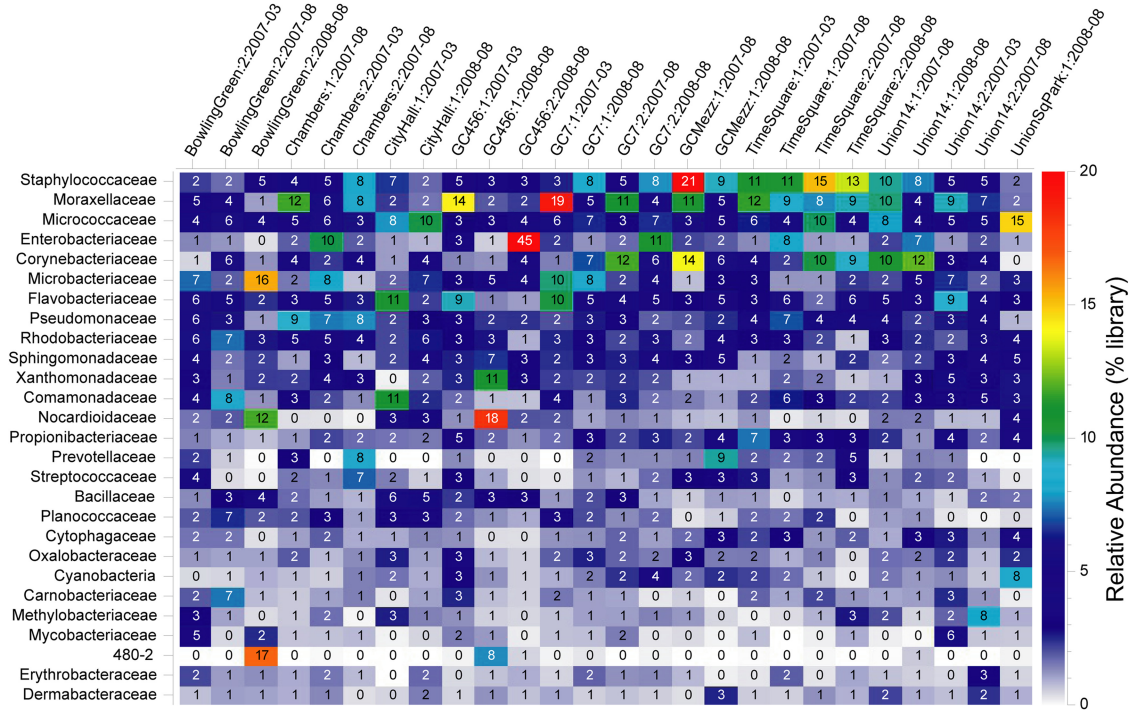

FIG 3 Abundances of bacterial family level aerosol taxa in NYC subway stations and Union Square Park (outside). Family level taxonomic categories with abundances of ≥1% in the particular sample are shown ranked in abundance from top to bottom. The sample naming convention is as follows: the station name, followed by the station sample number (samples were taken at both ends of platforms), followed by the date the sample was obtained. For complete taxonomic classifications, see Table S1 in the supplemental material. The percentage of each category is indicated by the value in each colored square of the heat map. The values do not add up to 100% because of rounding. Abbreviations: GC, Grand Central Station (train platforms 4, 5, and 6 and train platform 7 are indicated); GCMezz, Grand Central Station mezzanine; Union14, Union Square and 14th Street Station; UnionSqPark, Union Square Park (proxy for outdoors).

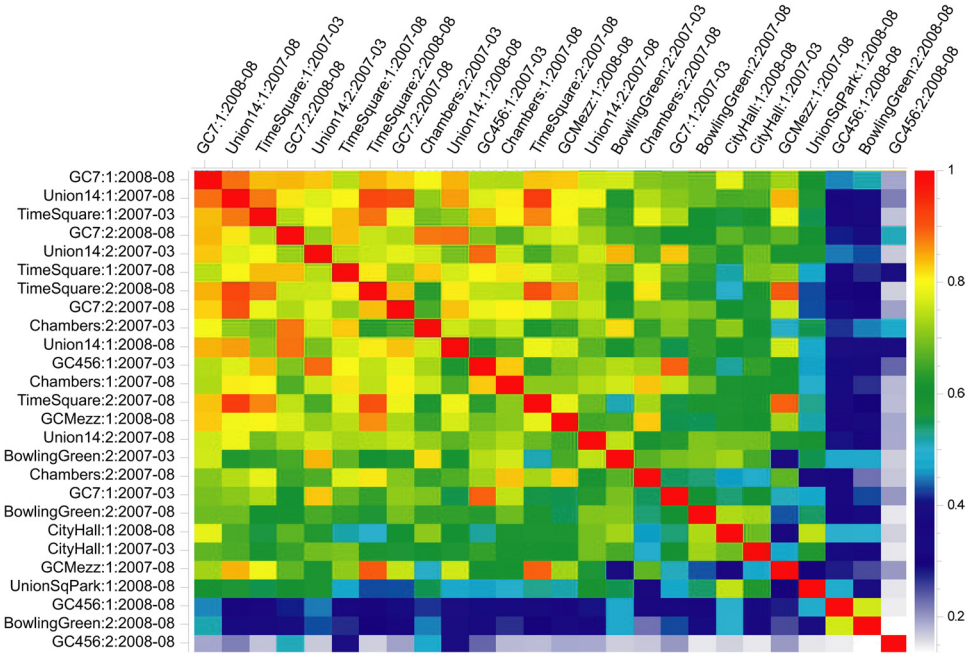

FIG 4 Morisita-Horn diversity comparison of aerosol taxa in NYC subway stations and Union Square Park (outside). Family level taxonomic classifications of pyrosequence data were used to compute the Morisita-Horn beta-diversity indexes for all sample pairs. Sample pairs with the most overlap in all taxonomic categories are shown as red squares, while sample pairs with the least overlap are shown as white squares. Columns are oriented from most similar (top) to least similar to other sites. Abbreviations: GC, Grand Central Station (train platforms 4, 5, and 6 and train platform 7 are indicated); GCMezz, Grand Central Station mezzanine; Union14, Union Square Park and 14th Street Station; UnionSqPark, Union Square Park (proxy for outdoors).

consistently higher levels than other taxa, but the distributions of the abundant taxa varied between different subway stations and times of sampling. For instance, taxonomic families that were particularly abundant in one or more samples occurred in other samples as well but in lesser relative abundance (Fig. 3).

We explored potential correlations in the distributions of the bacterial diversity between different subway sites by using standard ecological beta-diversity indices (i.e., Morisita-Horn and UniFrac) but identified few associations. As shown in Fig. 4, a Morisita-Horn analysis of similarities between the compositions of the different samples, microbiotas tended to be similar in general (M-H index, 0.6 to 0.8; identical communities would have an index of 1.0). The Morisita-Horn analysis does indicate a few pairs of stations with exceptionally high similarities (M-H index, >0.9), which in some instances can be rationalized by proximity. However, potential correlations such as station proximity tend to be inconsistent in subsequent samplings.

A few samples in Fig. 4 also are exceptionally different from the others. These can be explained largely by idiosyncratic abundances of particular phyla. For instance, dissimilarity between the GC456:1:2008-08 sample and other samples (Fig. 4; M-H index, <0.3) was driven by an unusual abundance of *Enterobacteriaceae* sequences (Fig. 4); dissimilarity of the BowlingGreen:1:2008-08 sample from others resulted from idiosyncratic abundances of *Microbacteriaceae*, *Nocardioidaceae*, and 480-2 (an environmental actinobacterial sequence clade with no named representative) in that sample (Fig. 3). Moreover, the somewhat unusual microbial compositions observed in the GC456:2:2008-08 and BowlingGreen:1:2008-08 samples was not reproduced in samples taken at the same sites but a year or more earlier (GC456, March 2007; BowlingGreen, August 2007), when the microbiology of those sites was generally similar to that of other stations sampled. In general, however, any apparent correlations have no particular and consistent explanations in terms of the sites or time points sampled. Samples taken at the same sites months or a year apart usually had no specific relationships; even samples taken on the same dates at opposite ends of particular platforms (e.g., samples TimesSquare:1:2007 and TimesSquare:2:2007 and samples GC456:1:2008-08 and GC456:2:2008-08) were no more correlated with each other than with other samples (M-H index, <0.9).

Sources of bacterial diversity. At low taxonomic levels, i.e., genus and species, the sequence diversity seen throughout the subway system was complex. However, most of the sequences fall into taxonomic groups that also contain described and named species (Fig. 3); 65 and 70% of the pyrosequences and Sanger sequences, respectively, were classifiable to the genus level. This is an unusually detailed taxonomic result with environmental samples, which typically yield many sequence types not in reference databases (33). The result indicates that the diversity we detected generally belongs to a relatively intensively documented portion of the bacterial SSU rRNA tree. Many of the sequences determined were nearly identical to the rRNA gene sequences of named species, and in some cases, the nature of those species points to the sources of the microorganisms encountered in the system.

In order to gain some further perspective on the sources of the sequences, we used BLAST analysis to compare the subway sequences to those of databases representing other environments. As general reference environments, we considered human skin, soil, and water (see Materials and Methods). Figure 5A shows comparisons of the subway data set with those environmental data sets at 95% sequence overlap and two levels: 95% sequence iden-

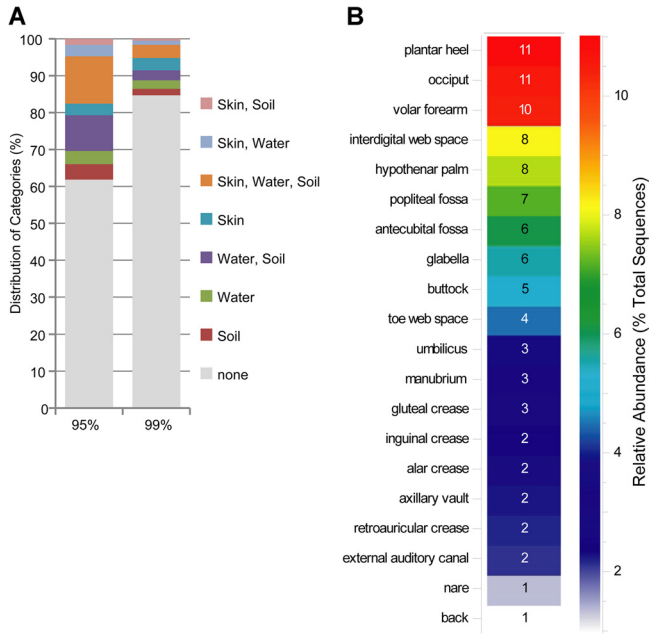

FIG 5 Comparison of aerosol DNA sequences from the NYC subway and Union Square Park with DNA sequences obtained from human skin, water, and soil. (A) Bacterial pyrosequence data were compared by BLAST to databases containing sequences from the HSM (27) and sequences obtained from the SILVA 108 ribosomal sequence database, and the source is indicated as either soil or water. A match between an aerosol sequence and the reference was identified whenever the pyrosequence matched a database sequence with at least 95% overlap and was at least (columns) 95% identical (genus or species level relationship) or at least 99% identical (close species relationship). Sequence matches were sorted into eight (non-mutually exclusive) categories. Sequences that did not match any of the three databases went into "none," and those that matched both human skin and soil sequences were assigned to "Skin, Soil," etc. (B) Potential human skin sources of aerosol DNA sequences from NYC subway stations were estimated by comparison of pyrosequence data by BLAST to a set of HSM databases created by random resampling of skin sites to the size of the smallest skin site sequence set (~3,800 sequences/site). A match between a pyrosequence and a skin site sequence was called when the pyrosequence overlapped the site sequence by at least 95% and was 99% identical. The percentages of these high-stringency matches are shown on a spectrum from red (>25% of pyrosequences match) to white (0% of pyrosequences match). Skin sites are defined by metadata for sequences deposited in the GenBank HSM database. The percentage of skin-related sequences in each category is indicated by the value in each colored square. The values do not add up to 100% because of rounding.

tity, approximately the genus level, and 99% identity, a stringent species level relationship. The results of the comparisons establish the general propensity of identifiable sequences to sort with both environmental (soil and water) and human skin microbiota, as indicated. Approximately 20% of the identifiable species are associated with human skin. Additional detail is provided in Fig. 5B, which indicates that sequences of identifiable human skin microbiota in the subway station air seem to be derived mainly from the feet, hands, arms, and head. For details of the body site representations for different stations, see Fig. S2 in the supplemental material.

One conspicuous example of putative human skin microbes in subway aerosols is the most abundant taxonomic family throughout the system (Fig. 3), *Staphylococcaceae*. Overall, 85% of the sequences representing *Staphylococcaceae* were members of the genus *Staphylococcus* in the SILVA 108 database. Most of the *Staphylococcus* species sequences observed were not affiliated with named species but rather represented a broad phylogenetic distribution within the genus. However, 475 of the ~3,000 *Staphylococcus* species sequences were identified (99% sequence identity) with 5 well-studied species, i.e., *S. epidermidis* (the most abundant by ~4-fold and the most abundant and prevalent commensal on the human integument), *S. hominis*, *S. cohnii*, *S. caprae*, and *S. haemolyticus*. These are all well-documented human commensal bacteria associated mainly with skin. *Staphylococcus* spp. were particularly abundant in areas of human habitation, for instance, the Grand Central Station mezzanine (GCMez:1:2007-08) and Times Square stations (Fig. 3). Similar human-related microbiological signatures were also represented in other of the family groups cited in Fig. 3. The *Micrococcaceae* sequence collection, for instance, includes *Kocuria* spp., *Micrococcus* spp., and others known as human skin microbiota.

Of course, many organisms related to known environmental microbes were also identifiable. For instance, ~20% of the *Micrococcaceae* sequences correspond to those of *Arthrobacter* spp., which are related to common soil organisms. The *Moraxellaceae* sequences, as a further example, include many *Acinetobacter* and *Psychrobacter* species sequences related to known soil and water microbes. Overall, as summarized in Fig. 5A, the bacterial diversity of the subway system aerosol that can be identified as to source was composed mainly of soil and water microbes with a significant overlay of human skin microorganisms. No pathogens beyond those associated with human commensal organisms were observed.

Eukaryotic diversity. The Sanger sequences based on universal primers provide the only perspective on aerosolized eukaryotic diversity detected in this study. Nearly all (96%) of the eukaryotic sequences corresponded to diverse fungi. The few eukaryotic rRNA sequences beyond fungal sequences included insect, plant, rodent, and a few protist sequences. Fungal sequences made up ~35 and 40% of the Sanger sequences of subway and outdoor rRNA genes, respectively (Fig. 2). These relatively high abundances of fungal rRNA genes compared to bacterial genes cannot be taken as cellular (or spore) frequencies relative to bacteria, however, because of the common occurrence of large numbers of rRNA genes per genome in fungi, typically 100 or more (34, 35). In contrast, bacteria are expected to have only a few rRNA genes (36). Consequently, the abundance of fungi relative to that of bacteria in cell numbers is probably overestimated in these sequence sets by 10- to 100-fold. The low frequency of fungus spores observed in direct counts was consistent with that expectation.

The taxonomic breakdown of the fungal sequences detected was complex, and as with the bacterial sequences, few corresponded exactly to named organisms (Fig. 6). All, however, fall into more or less well-described groups, including mushrooms (e.g., representatives of *Agaricomycetes*), mildews (e.g., members of *Eurotiales*), saprophytes and plant pathogens (e.g., *Capnodiales*), yeasts (e.g., *Saccharomycetales*), and wood-rotting fungi (most of the taxa listed in Fig. 6). No sequences of known human pathogens were present in the data sets. Although most of the fungal sequences were comparable between outside and subway samples, approximately 20% of the fungal sequences in the subway samples had no counterparts in the outside libraries (Fig. 6). Wood rot fungi were common among all of the sequences detected, so the increased fungal diversity among the subway sam-

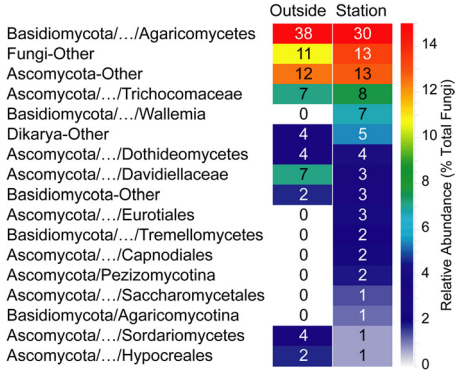

FIG 6 Comparison of aerosolized fungal DNA sequences from NYC subway stations and Union Square Park. Subway data were pooled by averaging the percent abundances of fungal sequences in each taxonomic category shown. The percentage of each category is indicated by the value in each colored square. The values do not add up to 100% because of rounding.

ples is likely due to the presence of the wooden track structure found throughout the system.

DISCUSSION

The subway environment, underground and away from light, might seem remote from our usual environment and potentially occupied by distinct or novel kinds of microorganisms. Rather than unusual, however, our survey finds that the microbiota encountered in the NYC subway is fairly mundane, essentially a mixture of outdoor air with an overlay of human-associated microorganisms typical of the skin. No significant evidence of pathogens or other organisms of concern was obtained, beyond what might be encountered in any human-occupied indoor setting. Thus, this survey provides the pre-event information necessary for surveillance activities for pathogens that might occur or be introduced into the system. The results also provide pre-event information necessary for interpretation of the microbiological consequences of the recent flooding of the NYC subway system during Hurricane Sandy in 2012.

Although subways might be considered confined environments, the similarity of subway air microbiota to that of outside air suggests that subway air significantly equilibrates with outside air on relatively short time scales. There is little or no local air conditioning in the NYC subway. Instead, air movement in the system is driven by passive train pumping, with air taken in and exhausted through street level ports, the NYC sidewalk grillwork. The general uniformity of microbial assemblages throughout the system indicates good air mixing, a testimony to the efficiency of the train-pumping process.

We have no information as to the viability of any of the organisms detected by gene sequences. Indeed, the direct cell counts reported here are ~100-fold higher than the CFU counts reported in other studies of subway aerosols (13), certainly indicating that most of what we identify is not cultured by standard methods. It is possible, perhaps likely, that many or most aerosolized organisms are nonviable because uncontrolled desiccation of bacteria can be lethal (12, 14, 37). Although bacterial and fungal spores are long-lived in a desiccated or partially desiccated condition, in the case of bacteria, we see little evidence of the specific selection of phylotypes known for spore formation. For instance, rRNA sequences representative of the spore-forming *Bacillaceae* family, common soil constituents, are only a minor constituent of the subway assemblage (~11% of the total bacteria, Fig. 3). We did not detect any potential bioterrorism agents, such as the spore-forming organism *Bacillus anthracis*. On the other hand, health-related concerns are not restricted to virulent microbes, since allergic and hypersensitive responses to bacteria and fungi do not require viable organisms.

No standing-crop aerosol microorganisms are known, so the microbes detected in this survey are not expected to be a "community" of interacting organisms. Rather, we consider the aerosol microbiota to be an assemblage of microbes derived from microenvironment-associated communities the air has contacted. Although most of the microorganisms are of unknown origin beyond vague descriptions such as "soil" or "water," human skin is likely to be a significant microenvironmental source of identifiable aerosolized microbes. This identification is possible because the microbial diversity associated with skin has been characterized intensively by culture and molecularly, and the rRNA sequence database of named skin microbes is large (27). Consequently, we are able to use a stringent criterion, 99% sequence identity, for the identification of subway microbes to the species level on the basis of sequence identity. (Most environmental studies use a more relaxed 97% identity rRNA sequence bin for species level taxonomic calls, which incorporates a much more diverse collection of organisms than our conservative estimate.) The 99% identity of rRNA gene sequences of many subway microbes with the corresponding sequences of specific human skin commensal microbes is strong evidence for our conclusion that a significant component of subway aerosol microorganisms is human derived. This is not surprising; many of the microorganisms in human-occupied indoor environments have long been identified with human skin microbiology (14, 38, 39).

Overall, approximately 5% of the subway aerosol sequences correspond specifically to human skin bacterial sequences (Fig. 5). Although our sampling campaign focused on occupied stations, similar results were seen with aerosols collected at an unused station (City Hall, Fig. 3). This indicates that human-associated microorganisms are dispersed throughout the system. This is perhaps surprising, considering the large spatial volume of hundreds of kilometers of subway, even in the face of considerable human traffic. Although shedding of skin flakes carrying microbes is often invoked as a source of human skin microbes, we suggest that convection driven by body temperature may also be a major factor. That is, we humans all have a body temperature of ~37°C, but when exposed to ambient air, we are surrounded by a lower temperature, typically ~20°C. Thus, we all continuously and actively emit a convective plume of warm air—carrying our skin microorganisms selectively (40).

Each sampling event and DNA library analysis was conducted only once, so we acknowledge that any specific variation in microbial contents between the samples might result from statistical flux or intrinsic sampling variation. Nonetheless, although each of the samples is a unique snapshot, the census results are broadly consistent in most samples and they collectively identify the microorganisms likely to be encountered in the NYC subway system. We see no consistent evidence of local pockets of specific diversity. Rather, the diversity observed at different sites seems to reflect random draws on a complex assemblage distributed throughout the system.

The development of "next-generation" sequencing technol-

ogy has dramatically changed the depth to which we can explore the natural microbial world, and the results have provided an entirely new perspective on environmental microbial diversity. Early molecular studies of natural microbial diversity tended to be limited, by technology and cost, to a few hundred sequences and described only a limited extent of environmental diversity; now, far deeper coverage is readily achieved. Sequence surveys such as the current one and many others have revealed that environmental microbiotas consist not simply of collections of specific microbes, "species," but rather complex collections of more or less closely related phylotypes, with little real demarcation between formal taxonomic grades such as genus and species. The results of high-volume sequencing of environmental microbiotas also begin to capture sequences that occur only rarely, a "rare biosphere" that sometimes contains far more phylogenetic diversity than that spanned by sequences of the more abundant organisms in the particular sampling (38). For instance, in this study, only four bacterial phyla were the major contributors to the subway aerosol microorganisms found. However, the inclusion of more rare sequences expanded the detection to ~12 phyla (see Table S2 in the supplemental material). Although not abundant, these rare sequences potentially represent significant environmental diversity and a significant contribution to the local pangenome pool.

ACKNOWLEDGMENTS

We thank Michael Metz, Mike Gemelli, Charles Burrus, and Frank Klimasz of the NYC Metropolitan Transit Authority for their assistance in conducting the sampling campaign. We also thank Kimberly Ross, Piret Koll, and Leah Feazel for assistance in sampling and Paula Olsiewski of the Alfred P. Sloan Foundation for encouragement and support.

This study was supported by grants from the Alfred P. Sloan Foundation (N.R.P.) and NIH (HG005964 [D.N.F.]).

N.R.P. and L.K.B. conceived and designed the experiments; L.K.B., K.L.P., and M.J.S. performed the experiments; C.E.R., D.N.F., J.K.H., N.R.P., and L.K.B. analyzed the data; and N.R.P., D.N.F., C.E.R., J.K.H., and L.K.B. wrote the paper.

REFERENCES

1. **Awad A.** Environmental study in subway metro stations in Cairo, Egypt. J. Occup. Health **44**:112–118.
2. **Cho JH, Hee Min K, Paik NW.** 2006. Temporal variation of airborne fungi concentrations and related factors in subway stations in Seoul, Korea. Int. J. Hyg. Environ. Health **209**:249–255.
3. **Jung H-J, Kim B, Malek MA, Koo YS, Jung JH, Son Y-S, Kim J-C, Kim H, Ro C-U.** 2012. Chemical speciation of size-segregated floor dusts and airborne magnetic particles collected at underground subway stations in Seoul, Korea. J. Hazard. Mater. **213–214**:331–340.
4. **Salma I, Posfai M, Kovacs K, Kuzmann E, Homonnay Z, Posta J.** 2009. Properties and sources of individual particles and some chemical species in the aerosol of a metropolitan underground railway station. Atmos. Environ. **43**:3460–3466.
5. **Nieuwenhuijsen M.** 2007. Levels of particulate air pollution, its elemental composition, determinants and health effects in metro systems. Atmos. Environ. **41**:7995–8006.
6. **Midander K, Elihn K, Wallén A, Belova L, Karlsson AK, Wallinder IO.** 2012. Characterisation of nano- and micron-sized airborne and collected subway particles, a multi-analytical approach. Sci. Total Environ. **427–428**:390–400.
7. **Chillrud SN, Epstein D, Ross JM, Sax SN, Pederson D, Spengler JD, Kinney PL.** 2004. Elevated airborne exposures of teenagers to manganese, chromium, and iron from steel dust and New York City's subway system. Environ. Sci. Technol. **38**:732–737.
8. **Chillrud SN, Grass D, Ross JM, Coulibaly D, Slavkovich V, Epstein D, Sax SN, Pederson D, Johnson D, Spengler JD, Kinney PL, Simpson HJ, Brandt-Rauf P.** 2005. Steel dust in the New York City subway system as a source of manganese, chromium, and iron exposures for transit workers. J. Urban Health **82**:33–42.
9. **Grass DS, Ross JM, Family F, Barbour J, James Simpson H, Coulibaly D, Hernandez J, Chen Y, Slavkovich V, Li Y, Graziano J, Santella RM, Brandt-Rauf P, Chillrud SN.** 2010. Airborne particulate metals in the New York City subway: a pilot study to assess the potential for health impacts. Environ. Res. **110**:1–11.
10. **Birenzvige A, Eversole J, Seaver M, Francesconi S, Valdes E, Kulaga H.** 2003. Aerosol characteristics in a subway environment. Aerosol Sci. Technol. **37**:210–220.
11. **Gilleberg S, Faull JL, Graeme-Cook KA.** 1998. A preliminary survey of aerial biocontaminants at six London Underground stations. Int. Biodeterior. Biodegradation **41**:149–152.
12. **Seino K, Takano T, Nakamura K, Watanabe M.** 2005. An evidential example of airborne bacteria in a crowded, underground public concourse in Tokyo. Atmos. Environ. **39**:337–341.
13. **Hwang S, Yoon C, Ryu K, Paik S, Cho J.** 2010. Assessment of airborne environmental bacteria and related factors in 25 underground railway stations in Seoul, Korea. Atmos. Environ. **44**:1658–1662.
14. **Dybwad M, Granum PE, Bruheim P, Blatny JM.** 2012. Characterization of airborne bacteria at an underground subway station. Appl. Environ. Microbiol. **78**:1917–1929.
15. **Pace NR.** 1997. A molecular view of microbial diversity and the biosphere. Science **276**:734–740.
16. **Kesavan JS, Schepers DR.** 2006. Characteristics and sampling efficiencies of Omni 3000 aerosol samplers. Edgewood Chemical Biological Center, Aberdeen, MD.
17. **Lane DJ, Pace B, Olsen GJ, Stahl DA, Sogin ML, Pace NR.** 1985. Rapid determination of 16S ribosomal RNA sequences for phylogenetic analyses. Proc. Natl. Acad. Sci. U. S. A. **82**:6955–6959.
18. **Hamady M, Walker JJ, Harris JK, Gold NJ, Knight R.** 2008. Error-correcting barcoded primers for pyrosequencing hundreds of samples in multiplex. Nat. Methods **5**:235–237.
19. **Harris JK, Sahl JW, Castoe TA, Wagner BD, Pollock DD, Spear JR.** 2010. Comparison of normalization methods for construction of large, multiplex amplicon pools for next-generation sequencing. Appl. Environ. Microbiol. **76**:3863–3868.
20. **Frank DN.** 2008. XplorSeq: a software environment for integrated management and phylogenetic analysis of metagenomic sequence data. BMC Bioinformatics **9**:420. doi:10.1186/1471-2105-9-420.
21. **Frank DN.** 2009. BARCRAWL and BARTAB: software tools for the design and implementation of barcoded primers for highly multiplexed DNA sequencing. BMC Bioinformatics **10**:362. doi:10.1186/1471-2105-10-362.
22. **Nawrocki EP, Kolbe DL, Eddy SR.** 2009. Infernal 1.0: inference of RNA alignments. Bioinformatics **25**:1335–1337.
23. **Haas BJ, Gevers D, Earl AM, Feldgarden M, Ward DV, Giannoukos G, Ciulla D, Tabbaa D, Highlander SK, Sodergren E, Methé B, DeSantis TZ, Petrosino JF, Knight R, Birren BW.** 2011. Chimeric 16S rRNA sequence formation and detection in Sanger and 454-pyrosequenced PCR amplicons. Genome Res. **21**:494–504.
24. **Li E, Hamm CM, Gulati AS, Sartor RB, Chen H, Wu X, Zhang T, Rohlf FJ, Zhu W, Gu C, Robertson CE, Pace NR, Boedeker EC, Harpaz N, Yuan J, Weinstock GM, Sodergren E, Frank DN.** 2012. Inflammatory bowel diseases phenotype, C. difficile and NOD2 genotype are associated with shifts in human ileum associated microbial composition. PLoS One **7**:e26284. doi:10.1371/journal.pone.0026284.
25. **Pruesse E, Quast C, Knittel K, Fuchs BM, Ludwig W, Peplies J, Glöckner FO.** 2007. SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. Nucleic Acids Res. **35**:7188–7196.
26. **Magurran AE.** Measuring biological diversity. Blackwell Publishing Company, Malden, MA.
27. **Grice EA, Kong HH, Conlan S, Deming CB, Davis J, Young AC, Bouffard GG, Blakesley RW, Murray PR, Green ED, Turner ML, Segre JA.** 2009. Topographical and temporal diversity of the human skin microbiome. Science **324**:1190–1192.
28. **Toivola M, Alm S, Reponen T, Kolari S, Nevalainen A.** 2002. Personal exposures and microenvironmental concentrations of particles and bioaerosols. J. Environ. Monit. **4**:166–174.
29. **Kujundzic E, Zander DA, Hernandez M, Angenent LT, Henderson DE, Miller SL.** 2005. Effects of ceiling-mounted HEPA-UV air filters on air-

borne bacteria concentrations in an indoor therapy pool building. J. Air Waste Manage. Assoc. **55**:210–218.
30. **Fabian M, Miller S, Reponen T, Hernandez M.** 2005. Ambient bioaerosol indices for indoor air quality assessments of flood reclamation. J. Aerosol Sci. **36**:763–783.
31. **Kuczynski J, Liu Z, Lozupone C, McDonald D, Fierer N, Knight R.** 2010. Microbial community resemblance methods differ in their ability to detect biologically relevant patterns. Nat. Methods **7**:813–819.
32. **Sogin ML, Morrison HG, Huber JA, Mark Welch D, Huse SM, Neal PR, Arrieta JM, Herndl GJ.** 2006. Microbial diversity in the deep sea and the underexplored "rare biosphere." Proc. Natl. Acad. Sci. U. S. A. **103**:12115–12120.
33. **Harris JK, Caporaso JG, Walker JJ, Spear JR, Gold NJ, Robertson CE, Hugenholtz P, Goodrich J, McDonald D, Knights D, Marshall P, Tufo H, Knight R, Pace NR.** 2013. Phylogenetic stratigraphy in the Guerrero Negro hypersaline microbial mat. ISME J. **7**:50–60.
34. **Ganley AR, Kobayashi T.** 2007. Highly efficient concerted evolution in the ribosomal DNA repeats: total rDNA repeat variation revealed by whole-genome shotgun sequence data. Genome Res. **17**:184–191.
35. **Maleszka R, Clark-Walker G.** 1990. Magnification of the rDNA cluster in Kluyveromyces lactis. Mol. Gen. Genet. **223**:342–344.
36. **Lee ZM, Bussema C, III, Schmidt TM.** 2009. rrnDB: documenting the number of rRNA and tRNA genes in bacteria and archaea. Nucleic Acids Res. **37**:D489–D493.
37. **Potts M.** 2001. Desiccation tolerance: a simple process? Trends Microbiol. **9**:553–559.
38. **Rintala H, Pitkäranta M, Toivola M, Paulin L, Nevalainen A.** 2008. Diversity and seasonal dynamics of bacterial community in indoor environment. BMC Microbiol. **8**:56. doi:10.1186/1471-2180-8-56.
39. **Hospodsky D, Qian J, Nazaroff WW, Yamamoto N, Bibby K, Rismani-Yazdi H, Peccia J.** 2012. Human occupancy as a source of indoor airborne bacteria. PLoS One **7**:e34867. doi:10.1371/journal.pone.0034867.
40. **Gao N, Niu J.** 2005. CFD study of the thermal environment around a human body: a review. Indoor Built Environ. **14**:5–16.